B636- Next Generation Genomic Data Analytics
Fall 2013
Class room: IT271, 3 Credit Hours

Instructor: Sarath Chandra Janga, PhD, Assistant Professor, Bioinformatics
Office Address: 719 Indiana Avenue, Room WK 309 (Walker Plaza Building)
Office Phone: 317-278-4147
Office Hours: Tue, Thu, 11:00am-12:00pm
Email Address: jangalab@iupui.edu

The Mission of IUPUI is to provide for its constituents excellence in
- **Teaching and Learning**
- **Research, Scholarship, and Creative Activity**
- **Civic Engagement**

With each of these core activities characterized by
- **Collaboration within and across disciplines and with the community**
- **A commitment to ensuring diversity, and**
- **Pursuit of best practices**

IUPUI's mission is derived from and aligned with the principal components – Communities of Learning, Responsibilities of Excellence, Accountability and Best Practices – of Indiana University's Strategic Directions Charter.

Statement of Values

IUPUI values the commitment of students to learning; of faculty to the highest standards of teaching, scholarship, and service; and of staff to the highest standards of service. IUPUI recognizes students as partners in learning. IUPUI values the opportunities afforded by its location in Indiana's capital city and is committed to serving the needs of its community. Thus, IUPUI students, faculty, and staff are involved in the community; both to provide educational programs and patient care and to apply learning to community needs through service. As a leader in fostering collaborative relationships, IUPUI values collegiality, cooperation, creativity, innovation, and entrepreneurship as well as honesty, integrity, and support for open inquiry and dissemination of findings. IUPUI is committed to the personal and professional development of its students, faculty, and staff and to continuous improvement of its programs and services.

Electronic course site:
- oncourse.iu.edu
- distribution of homework assignments and class notes
- should be accessible if you have IU ID and registered for the course

Course website: http://www.iupui.edu/~jangalab/B636-fall-2014.php
- Syllabus, outline of lectures, course calendar and suggested reading
- Final projects and presentation schedules

Lectures location & time:
- IT 271, Tuesday 3:00pm-5:40pm
- May have some guest lectures, not necessarily in the same room and time

Textbook:
- class notes will be distributed via course website and the following books are recommended (for course assignments, exercises and projects)

Reading (referred recommended books for prerequisites in programming):
- C. Gibas & P. Jambeck, "Developing Bioinformatics Computer Skills" O'Reilly, 2001.
- J. Tisdall, "Beginning Perl for Bioinformatics," O'Reilly 2001.
- C. Bessant, I. Shadforth & D. Oakley, "Building Bioinformatics Solutions: with Perl, R and MySQL", Oxford University Press 2009
- B.S. Everitt & T. Hothorn, "A Handbook of Statistical Analyses Using R", Chapman & Hall/CRC 2006

Homework & Assignments:
- monthly homeworks (up to five programming assignments total)
- midterm (will be a take-home assignment)
- final project (read a series of papers, tackle a research problem, present the papers, demonstrate the project and write-up a short report – work in groups up to a maximum of two members per group)

Grading:
- Homeworks/Assignments (50%), Midterm (20%), Final presentation and project (30%)

Course Description and Outcomes:

In this course, we will cover the advanced concepts of a huge variety of genomic sequencing datasets emerging in the post-genomic era from a number of sequencing platforms. This will be achieved by giving a molecular biology/technological background of how the data is generated to motivate computational needs/tasks for analysis. Basics of programming are essential as the instructor will be introducing a wide range of existing tools and approaches to analyze these sequencing datasets and will be anticipating the students to either use or integrate the approaches to address specific questions in the assignments. At the end of the course, you should be able to describe solutions (preferably elegant) to analyze a wide range of clinical and biomedical sequencing datasets to provide sound hypothesis and predictions from massive amounts of data.

This course is an advanced genomics course aimed at giving a good foundation in sequencing technologies and their data analytics by introducing various existing implementations to suit specific tasks. Naturally expects knowledge of UNIX based administration, PERL programming, MySQL database management (optional) and R statistical analysis in order to work with the datasets.

The instructor will give introductions to each of the various types of sequencing datasets and commonly used applications in analyzing them in the first 10 weeks. Then

the students will be asked to present recent articles published in the last 4 years (each student has to present couple of papers), present a project work on a particular theme/problem and submission of the project report.

TENTATIVE WEEKLY SCHEDULE (subjected to change)

Week	Topics Covered	Reading	Assignments
1	Outline, syllabus, evaluation procedures and primer on molecular biology and sequencing platforms.		A1: Assemble a bacterial genome given short-read data
2	Next generation DNA sequencing platforms, quality control, alignment (problems and solutions), Assembly of genomes (computational approaches)		
3	RNA-sequencing protocols, data storage and file formats, various aligners, expression quantification methods, differential expression analysis		A2: Given two GEO fastq datasets, identify and analyse the differentially expressed transcripts.
4	Microbial DNA and RNA sequencing, computational approaches for assembly, quantification and detection of operons and transcription units. Introduction to Genome browsers.		A3.Given a fastq raw reads for a bacterial genome, align, quantify and construct the transcriptome map of the genome.
5	Metagenomics/ Metatranscriptomics: Computational methods to assemble, taxonomic mapping, analyze and study microbial abundance, interactions and pathways.		
6	ChIP-sequencing and related approaches. Computational approaches and analysis. Concept of peak calling and benchmarking studies illustrating the variations between methods.		A4: Analyze raw ChIP-seq data for a TF across cell types from encode data and analyze the differences.
7	CLIP-sequencing and related approaches for building post-transcriptional networks. Computational approaches and analysis.		Mid-term quiz: Take home assignment and/or quiz.
8	Misc datasets: Small RNA sequencing/ protein occupancy profiling/ polyA sequencing		
9	Concept of genetic and somatic variation. Variant calling approaches. Issues and Considerations.		A5: Identify and delineate the germline and somatic SNPs from a series of samples taken from population sequencing project.
10	Expression Quantitative Trait Loci identification/ Ethical considerations of personalized sequencing projects.		

11	Project presentations		
12	Project presentations		
13	Project presentations		Project reports due for submission.
14	Project presentations		

Core Competencies: Back ground of molecular biology and basic biochemistry is expected although several of the required concepts will be revised throughput the course. Programming skills are expected and are a prerequisite for this course. If you are uncertain course I573 (programming for chemical and life sciences) is recommended before taking this course.

Software used: Unix operating system, Perl, R and SQL (Either Oracle and Mysql)

Expectations/Guidelines/Policies:

Late Policy: You are expected to submit assignments on the due dates unless you have a medical condition which warrants it (supported by medical evidence).

Attendance: Attendance and punctuality is expected and may be encouraged/rewarded with bonus points up to 5% which will be discussed in the first lecture.

Cell phone: Cell phones are not allowed in the class unless they are switched off. Failure to comply with this policy could negatively impact your attendance scores.

Browsing: Laptops/desktops should be strictly used for course related content and chatting is banned during the lecture.

Copying policy: Help and collaboration between yourselves is encouraged for working on your assignments but copying from one another and from the internet is strictly prohibited. *Note that it is easy to identify code which is copied.*

Grading Scale
 A+ (93-100%), A (88-92) and B (84-87%) **[Tentative]**

Course Evaluation

Homework/ Assignments (50%)
–A total of five programming assignments (50%)
–At least one assignment each on major concepts such as DNA-seq assembly, RNA-seq data analysis, ChIP-sequencing, CLIP-sequencing data etc. (available from on course website)
–Strictly no copying of code from others and all the code should be functional to be awarded any grade.

Mid-term project evaluation (20%)
–A project on developing a pipeline for analyzing a novel type of sequencing dataset will be posed by the instructor and the students are expected to develop code to address the problem (see oncourse website)
–The developed system should be submitted as a short report demonstrating its functionality and the conclusions derived from it for the chosen dataset.
–Alternatively or in addition to the programming task above, instructor might decide to take a mid-term quiz on the course content taught.

Final project (30%)
–Will be evaluated for 30% of the total points
–Will comprise of a project for which you can work in groups of up to 2 students
–Involves reading a series of papers on sequencing technologies and datasets, tackling a research problem, presenting the read papers, demonstrating the project developed based on the papers in the class and write-up of a short report on the project.
–Presentation of the project would be towards the end of the semester and will be evaluated as a group.
–Paper presentation would weigh 15% and project demonstration/report would be for 15%.
–Each group will present a different set of papers and project to work on.
–Any questions regarding the projects should be directed to jangalab@iupui.edu

General rules/notes:

"Children are not permitted to attend class with parents, guardians, or childcare providers. This conduct has the effect of unreasonably interfering with an individual's work or academic performance creating an offensive learning environment."

"A student must not violate course rules as contained in a course syllabus, which are rationally related to the content of the course or to the enhancement of the learning process in the course." *[Code of Student Rights, Responsibilities, and Conduct, page 29]*